ON THE VARIANCE FUNCTION OF THE MEAN OF A STATIONARY STOCHASTIC PROCESS

By

C. Asok

Maharashtra Association for Cultivation of Science
Pune

(Received : December, 1981)

1. Introduction

In Nutrition Science, it has been traditionally assumed that energy and protein requirements are fixed in an adult individual, maintaining body weight (FAO/WHO [3]) in the sense that day-to-day fluctuations in requirement are random in character arising from errors of measurement. The model underlying this assumption can be described as :

$$R_t = R + e_t \qquad \ldots(1)$$

where R_t represents the requirement on the t^{th} day, R is the true unknown value and e_t is the measurement error assumed to be negligible as compared to R.

If the model as described in (1) is valid, it may be concluded that the influence of variation in daily requirement can be eliminated altogether by averaging requirement over a number of days as necessary to provide what is called a habitual requirement. However, according to Sukhatme [5] and Sukhatme and Margen [6], requirement of an individual is not constant over time as assumed. On the other hand it varies from day to day and the variations in requirement are of a probabilistic kind and can often approximately be represented by a suitable stationary stochastic process such as (*i*) Autoregressive, (*ii*) Moving Average or (*iii*) Mixed autoregressive, moving average (Paranjpe [4], Tilve [8]). In the light of this new finding one needs to re-examine the question viz., whether the influence of variation in daily requirement can be eliminated altogether by averaging requirement over a number of days as necessary to provide what is called a habitual requirement. Table 1 compares the standard deviation of the daily intake with the standard deviation of of the weekly means on a subject of the experimental study reported by Acheson *et al.* [1]. The table also includes the standard deviation obtained from 2, 3 and 4 weekly means. This table is reproduced from Sukhatme and Margen [7].

TABLE 1

Standard deviations of daily intake and expenditure in Kcal/day compared with those of daily means based on 1, 2, 3, and 4 weekly periods.

	Daily	*1-week*	*2-weeks*	*3-weeks*	*4-weeks*
Intake	646	370	286	259	243
Expenditure	725	441	303	262	233

It can be observed from Table 1 that the standard deviation calculated from weekly means is still reduced, even if it is much larger than would be expected if the successive observations were independent. This observation leads us to the question of examining the behaviour of the variance function of the mean taken over n days of each of the three stationary stochastic processes viz. $AR(1)$, $MA(1)$, and $ARMA$ (1, 1) which are found to be approximate representation of the requirements.

2. Behaviour of the Variance Function of the Mean of a Stationary Stochastic Process

At first we will consider the case of the first order autoregressive process.

(*a*) *The first-order autoregressive process*

The first-order autoregressive $AR(1)$ process is represented by

$$Z_t=\phi_1 Z_{t-1}+a_t \qquad \dots(2)$$

where $\{Z_t\}$ is the process under consideration and $\{a_t\}$ is the white noise process which consists of a sequence of uncorrelated random variables with mean zero and variance σ_a^2. It is assumed without loss of generality that the mean of the process Z_t is zero. Otherwise one has to replace Z_t in (2) by $\tilde{Z}_t$ where $\tilde{Z}_t=Z_t-\mu$ is the deviation of the process from its mean μ.

The $AR(1)$ process represented by (2) is stationary

if $$|\phi| < 1 \qquad \ldots(3)$$

Autocorrelation at lag k of stationary $AR(1)$ process is

$$\rho_k = \phi_1^k, \qquad \ldots(4)$$

and variance of the process is

$$\sigma_z^2 = \frac{\sigma_a^2}{(1-\phi_1^2)} \qquad \ldots(5)$$

If γ_k denotes the autocovariance at lag k of the stationary $AR(1)$ process then

$$\gamma_k = \rho_k \, \sigma_z^2 \qquad \ldots(6)$$

If $$\bar{z}_n = \frac{1}{n} \sum_{t=1}^{n} z_t \qquad \ldots(7)$$

denotes the sample mean, then its variance can be expressed as

$$V(\bar{z}_n) = V \frac{1}{n} \sum_{t=1}^{n} z_t$$

$$= \frac{1}{n^2} \sum_{t=1}^{n} V(z_t) + \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Cov}\,(z_i, z_j)$$

$$= \frac{\gamma_o}{n} + \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \gamma_{j-i} \qquad \ldots(8)$$

$$= \frac{\gamma_o}{n} + \frac{2\gamma_o}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \rho_{j-i}$$

$$= \frac{\gamma_o}{n} + \frac{2\gamma_o}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \phi_1^{j-i}$$

$$= \frac{\gamma_o}{n} + \frac{2\gamma_o}{n^2} \cdot \frac{\phi_1}{1-\phi_1} \sum_{i=1}^{n-1} (1-\phi_1^{n-i})$$

$$=\frac{\gamma_o}{n}+\frac{2\gamma_o}{n^2}\cdot\frac{\phi_1}{1-\phi_1}\left[(n-1)-\frac{\phi_1(1-\phi^{n-1})}{1-\phi_1}\right]$$

$$=\frac{\gamma_o}{n}\left[\frac{1+\phi_1}{1-\phi_1}-\frac{2\phi_1(1-\phi_1^n)}{n(1-\phi_1)^2}\right]$$

$$=\frac{\gamma_o}{n}\cdot\lambda_n, \text{ say}$$

where $\lambda_n=\dfrac{1+\phi_1}{1-\phi_1}-\dfrac{2\phi_1(1-\phi_1^n)}{n(1-\phi_1)^2}$

As the process is stationary, the condition $|\phi_1|<1$ given in (3) implies that $\phi_1^n\to 0$ as $n\to\infty$.

$\therefore$ As $n\to\infty$, $\lambda_n\to\dfrac{1+\phi_1}{1-\phi_1}$, a finite constant and hence

$$V(\bar{z}_n)=\frac{\gamma_o}{n}\cdot\lambda_n\to 0 \text{ as } n\to\infty$$

(*b*) *The first-order moving average process*

The first-order moving average $MA(1)$ process is represented by

$$Z_t=a_t-\theta_1 a_{t-1} \qquad \ldots(9)$$

where Z_t and a_t are the same as before.

The process is stationary for all values of θ_1 and invertible if

$$|\theta_1|<1 \qquad \ldots(10)$$

Variance of the process is given by

$$V(\bar{z}_n)=V\left(\frac{1}{n}\sum_{t=1}^{n} z_t\right)$$

$$=V\left[\frac{1}{n^2}\{(a_1-\theta_1 a_o)+(a_2-\theta_1 a_1)+\ldots+(a_n-\theta_1 a_{n-1})\}\right]$$

$$=\frac{1}{n^2}V[-\theta_1 a_o+(1-\theta_1)(a_1+a_2+\ldots+a_{n-1})+a_n]$$

$$=\frac{1}{n^2}[(1+\theta_1^2)+(n-1)(1-\theta_1)^2]\,\sigma_a^2$$

$$=\frac{(1+\theta_1^2)\,\sigma_a^2}{n}\cdot\left[\frac{(1-\theta_1)^2}{1+\theta_1^2}+\frac{1}{n}\cdot\frac{2\theta_1}{(1+\theta_1^2)}\right]$$

$$=\frac{\gamma_o}{n}\,\lambda_n,\text{ say}$$

where $$\lambda_n=\frac{(1-\theta_1)^2}{(1+\theta_1^2)}+\frac{1}{n}\cdot\frac{2\theta_1}{(1+\theta_1^2)}$$

As $n\to\infty$, $\lambda_n\to\frac{(1-\theta_1)^2}{(1+\theta_1^2)}$, a finite constant

and hence $V(\bar{z}_n)=\frac{\gamma_o}{n}\cdot\lambda_n\to o$ as $n\to\infty$

In fact the same can be shown to be true in general for a Moving average MA (q) process of order q, given by

$$Z_t=a_t-\theta_1\,a_{t-1}\,\ldots-\theta_q a_{t-q} \qquad \ldots(11)$$

Variance of the mean of the process is

$$V(\bar{z}_n)=\frac{\gamma_o}{n}+\frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\gamma_{j-i}$$

$$=\frac{\gamma_o}{n}+\frac{2}{n^2}\sum_{k=1}^{q}(n-k)\,\gamma_k$$

$$=\frac{\gamma_o}{n}\left[1+\frac{2}{n}\sum_{k=1}^{q}(n-k)\,\rho_k\right]$$

$$=\frac{r_o}{n}\cdot\lambda_n\text{ say}$$

where $$\lambda_n=1+\frac{2}{n}\sum_{k=1}^{q}(n-k)\,\rho_k$$

$$=1+2\sum_{k=1}^{q}\left(1-\frac{k}{n}\right)\rho_k$$

$$=1+2\sum_{k=1}^{q}\rho_k-\frac{2}{n}\sum_{k=1}^{q}k\rho_k$$

As $n\to\infty$, $\lambda_n\to 1+2\sum_{k=1}^{q}\rho_k=\frac{(1-\theta_1-\theta_2-\ldots-\theta_q)^2}{(1+\theta_1^2+\theta_2^2+\ldots+\theta_q^2)}$,

a finite constant.

$$\therefore\qquad V(\bar{z}_n)=\frac{\gamma_o}{n}\,.\,\lambda_n\to o \text{ as } n\to\infty$$

(*c*) *The first-order autoregressive-first order moving average ARMA* (1,1) *process*

The *ARMA* (1,1) process is

$$Z_t-\phi_1 Z_{t-1}=a_t-\theta_1 a_{t-1} \qquad \ldots(12)$$

The process is stationary if

$$|\phi_1|<1 \qquad \ldots(13)$$

and invertible if

$$|\theta_1|<1 \qquad \ldots(14)$$

The autocovariance function of the process is

$$\sigma_z^2=\gamma_o=\frac{1+\theta_1^2-2\phi_1\theta_1}{1-\phi_1^2}\sigma_a^2 \qquad \ldots(15)$$

$$\gamma_1=\frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2}\sigma_a^2 \qquad \ldots(16)$$

$$\text{and } \gamma_k=\phi_1\gamma_{k-1},\ k\geqslant 2 \qquad \ldots(17)$$

Variance of the mean of the process is given by

$$V(\bar{z}_n)=\frac{\gamma_o}{n}+\frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\gamma_{j-i}$$

$$=\frac{\gamma_o}{n}+\frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_1^{j-i-1}\ \gamma_1$$

$$=\frac{\gamma_o}{n}+\frac{2\gamma_1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_1^{j-i-1}$$

$$=\frac{\gamma_o}{n}+\frac{2\gamma_1}{n^2(1-\phi_1)}\sum_{i=1}^{n-1}(1-\phi_1^{n-i})$$

$$=\frac{\gamma_o}{n}+\frac{2\gamma_1}{n^2(1-\phi_1)}\left[(n-1)-\frac{\phi_1(1-\phi_1^{n-1})}{(1-\phi_1)}\right]$$

$$=\frac{\gamma_o}{n}+\frac{2\gamma_o\,\rho_1}{n^2(1-\phi_1)}\left[(n-1)-\frac{\phi_1(1-\phi_1^{n-1})}{(1-\phi_1)}\right]$$

$$=\frac{\gamma_o}{n}.\ \lambda_n,\ \text{say}$$

where

$$\lambda_n=1+\frac{2(1-\phi_1\theta_1)(\phi_1-\theta_1)}{n(1-\phi_1)(1+\theta_1^2-2\phi_1\theta_1)}\left[(n-1)-\frac{\phi_1(1-\phi_1^{n-1})}{(1-\phi_1)}\right]$$

$$=1+\frac{2(1-\phi_1\theta_1)(\phi_1-\theta_1)}{(1-\phi_1)(1+\theta_1^2-2\phi_1\theta_1)}-\frac{2(1-\phi_1\theta_1)(\phi_1-\theta_1)}{n(1-\phi_1^2)^2(1+\theta_1^2-2\phi_1\theta_1)}$$

$$+\frac{2(1-\phi_1\theta_1)(\phi_1-\theta_1)}{n(1-\phi_1)^2(1+\theta_1^2-2\phi_1\theta_1)}.\ \phi_1^n$$

Since $|\phi_1|<1$ from (13), $\phi_1^n\to o$ as $n\to\infty$ and hence

$$\lambda_n\to 1+\frac{2(1-\phi_1\theta_1)(\phi_1-\theta_1)}{(1-\phi_1)(1+\theta_1^2-2\phi_1\theta_1)}=1+\frac{2\theta_1}{1-\phi_1},\ \text{a finite}$$

constant

$$\therefore\qquad V(\bar{z}_n)=\frac{\gamma_o}{n}.\ \lambda_n\to o\ \text{as}\ n\to\infty$$

3. *Discussion*

Thus for all the three stationary stochastic processes, viz., *AR*(1), *MA*(1), and *ARMA* (1,1), variance of the mean taken over n days approaches zero as n approches infinity. In the light of this result and in the context of the topic under discussion, one important point to be observed is well brought out by Table 2 which shows the values of $\frac{\lambda_n}{n}$ for different values of ϕ_1 and n in the case of a stationary *AR*(1) process. This table is reproduced from Sukhatme and Margen [6].

TABLE 2

Variance of the mean of n values with unit variance when successive observations follow a stationary AR(1) process

$n \phi_1$	0.0	0.50	0.66	0.80
1	1.00	1.00	1.00	1.00
3	0.33	0.61	0.73	0.83
5	0.20	0.45	0.58	0.72
7	0.14	0.35	0.48	0.58

From Table 2, it can be seen that ϕ_1 is of the order of 0:6 to 0.7 variance of the estimated requirement even when based on 7 day average will be three times as large as when based on the assumption of independence between successive observations. One could counter this argument by saying that it is more important to observe the fact that the variance of the mean will invariably decrease as the number of days over which the requirement is averaged increases and hence, one should be able to eliminate altogether the influence of variation in daily requirement by averaging requirement over a longer period such as 3 or 4 weeks, if not over a short period as one week, as necessary to provide what is called a habitual requirement. This argument would perhaps be valid if the stationary autoregressive process of order 1 (or any of the other two stationary processes) were an exact description of the observed phenomenon. *AR*(1) process is considered more to illustrate the regulatory character of the energy (Nitrogen) balance than as an exact description of the phenomenon represented by the observed energy (nitrogen) balance series. The point that $AR(1)$ process is not an exact description of the phenomenon can be observed from Table 3 where the variance of an individual's mean energy intake and balance based on n successive days are presented for Edholm's series on energy balance (Edholm *et. al.*, [2]). This table is reproduced from Sukhatme and Margen [7].

TABLE 3

Variance of an individual's Mean Energy Intake and balance based on n successive days (expressed as proportion of unit variance for n=1)

Period in days	*Observed values for*	
	intake	*Balance*
1	1.00	1.00
2	0.62	0.53
3	0.44	0.38
4	0.45	0.37
5	0.37	0.27
6	0.31	0.21
7	0.31	0.22

From Table 3 one can conclude that the variance gets stabilized as we increase the number of observations over which mean is taken. From this it appears that the stationary stochastic processes of the three types considered here cannot be exact description of the phenomenon represented by the observed series but only rough approximations. It appears that an additional interaction term should be brought in into the model which keeps the variance constant. This interaction term represents the interaction of the specialized environment with the genetic component. Longterm series produced under controlled conditions are not available to allow this further refinement of the model. This calls for the need of conducting well designed metabolic experiments under controlled conditions.

A consequence of the stabilization of variance is that the influence of variation in daily requirement cannot be eliminated altogother by averaging requirement over a number of days as necessary to provide what is called a habitual requirement. From this it follows that one has to define nutritional deficiency as a failure of the process to be in statistical control, rather than as a situation in which the observed intake falls short of the individual's true requirement.

REFERENCES

[1] Acheson, K.J., Campbell, I.T., Edholm, O.G., Miller D.S. and Stock M.J. (1980) : A longitudinal study of body weight and body fat changes in Antarctica *Am. J. Clin. Nutr.*

[2] Edholm, O. G., Adam J.M., Healy, M.J.R., Wolff, H.S., Goldsmith, R. and Best, T.W. (1970). : Food intake and energy expenditure of army recruits. *Brit. J. Nutr.* **24,** 1091,

[3] FAO/WHO Expert Committee Report (1973). : Energy and Protein Requirements. WHO Technical Report Series No. 522 (FAO Nutrition Meetings. Report Series No. 52).

[4] Paranjpe, S.A. (1980) : Time series analysis of daily dietary intake and nutrient output. Unpublished Ph. D. Thesis, Department of Biometry, Maharashtra Association for the Cultivation of Science and the University of Pune, Pune.

[5] Sukhatme, P.V. (1974) : The protein problem, its size and nature. *J. Roy. Statist. Soc.* **137, 166.**

[6] Sukhatme, P.V, and Margen, S. (1978). : Models for protein deficiency. *Am. J. Clin. Nutr.* **31,** 1237.

[7] Sukhatme, P.V. and Margen, S. (1981) : Auto-Regulatory homeostatic nature of energy balance. Paper read at the Summer Institute on "Newer Concepts of Nutrition and Health and their Implications for Social Policy held at Maharashtra Association for the Cultivation of Science, Pune.

[8] Tilve, S.G. (1979) : Studies in variation in intake and balance. Unpublished Ph. D. thesis, Department of Biometry, Maharashtra Association for the Cultivation of Science and the University of Pune.